STIC-ILL

From:	Hunt, Jennifer
Sent:	Tuesday, July 03, 2001 5:25 PM
To:	STIC-ILL
Subject:	Ref for 09/234,208

Please send me the following references ASAP.

Yamamoto et al., Nature 319: 230-234 1986

Scott et al., Mol. Cell Biology, 13:2247-2257, 1993

Petch et al., Mol. Cell Biology, 10:2973-2982, 1990

Basu et al, Mol. Cell Biology, 9: 671-677 1989

Flickinger et al., Mol. Cell Biology, 12: 883-893 1992

Thanks,

Jennifer Hunt
Patent Examiner, Art Unit 1642
CM1-8D06
(703)308-7548

17537-10 from the NCI and by a grant from the American Business Cancer Research Foundation. M.-C.H. was supported by a postdoctoral fellowship from the Cancer Research Institute. R.A.W. is an American Cancer Society Research Professor.

Received 30 August; accepted 15 November 1985.

1. Schubert, D. *et al. Nature* **249**, 224–227 (1974).
2. Padhy, L. C. *et al. Cell* **28**, 865–871 (1982).
3. Schechter, A. L. *et al. Nature* **312**, 513–516 (1984).
4. Schechter, A. L. *et al. Science* **229**, 976–978 (1985).
5. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
6. Mulligan, R. C. & Berg, P. *Science* **209**, 1422–1427 (1980).
7. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–580 (1980).
8. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
9. Biggin, M. D., Gibson, T. J. & Hong, G. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963–3965 (1983).
10. Blobel, G. *et al. Symp. Soc. exp. Biol.* **33**, 9–36 (1979).
11. Ullrich, A. *et al. Nature* **309**, 418–425 (1984).
12. Ullrich, A. *et al. Nature* **313**, 756–761 (1985).
13. Ebina, Y. *et al. Cell* **40**, 747–758 (1985).
14. Bishop, J. M. *A. Rev. Biochem.* **52**, 301–354 (1983).
15. Downward, J., Parker, P. & Waterfield, M. D. *Nature* **311**, 483–485 (1984).
16. Hunter, T., Ling, N. & Cooper, J. A. *Nature* **311**, 480–483 (1984).
17. Friedman, B. *et al. Proc. natn. Acad. Sci. U.S.A.* 3034–3038 (1984).
18. Davis, R. J. & Czech, M. P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4080–4084 (1985).
19. Hunter, T. *Trends biochem. Sci.* **10**, 275–280 (1985).
20. Doolittle, R. F. *et al. Science* **221**, 275–277 (1983).
21. Waterfield, M. D. *et al. Nature* **304**, 35–39 (1983).
22. Downward, J. *et al. Nature* **307**, 521–527 (1984).
23. Scherr, C. J. *et al. Cell* **41**, 665–676 (1985).
24. Neckameyer, W. S. & Wang, L. H. *J. Virol.* **53**, 879–884 (1985).
25. Yamamoto, T. *et al. Cell* **39**, 27–38 (1984).
26. Yamamoto, T. *et al. Nature* **319**, 230–234 (1986).

Similarity of protein encoded by the human c-*erb-B*-2 gene to epidermal growth factor receptor

Tadashi Yamamoto, Shuntaro Ikawa, Tetsu Akiyama*, Kentaro Semba, Nobuo Nomura†, Nobuyuki Miyajima, Toshiyuki Saito & Kumao Toyoshima

Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
* Meiji College of Pharmacy, 1-35-23, Nozawa, Setagaya-ku, Tokyo 154, Japan
† Nippon Medical School at Sakuragi, 1-10-19, Uenosakuragi, Taito-ku, Tokyo 110, Japan

A novel v-*erb-B*-related gene, c-*erb-B*-2, which has been identified in the human genome[1,2], maps to human chromosome 17 at q21 (ref. 40), and seems to encode a polypeptide with a kinase domain that is highly homologous with, but distinct from, that of the epidermal growth factor (EGF) receptor[1]. The c-*erb-B*-2 gene is conserved in vertebrates and it has been suggested[1] that the *neu* gene, detected in a series of rat neuro/glioblastomas[3], is, in fact, the rat c-*erb-B*-2 gene. Amplification of the c-*erb-B*-2 gene in a salivary adenocarcinoma and a gastric cancer cell line MKN-7 suggests that its over-expression is sometimes involved in the neoplastic process. To determine the nature of the c-*erb-B*-2 protein, we have now molecularly cloned complementary DNA for c-*erb-B*-2 messenger RNA prepared from MKN-7 cells. Its sequence shows that the c-*erb-B*-2 gene encodes a possible receptor protein and allows an analysis of the similarity of the protein to the EGF receptor and the *neu* product. As a consequence of chromosomal aberration in MKN-7 cells, a 4.6-kilobase (kb) normal transcript and a truncated 2.3-kb transcript of c-*erb-B*-2 are synthesized at elevated levels. The latter transcript presumably encodes only the extracellular domain of the putative receptor.

Poly(A)$^+$ RNA was prepared from MKN-7 cells, a gastric cancer cell line in which the c-*erb-B*-2 gene is amplified[40] and over-expressed (see below). A cDNA library constructed from the MKN-7 mRNA was screened initially with a 440-base pair (bp) KX DNA fragment prepared from a c-*erb-B*-2 gen clone and then with cDNA probes as described in Fig. 1 leg Among 113 positive clones, pCER204 carried the longest i of 4.0 kb, whereas the c-*erb-B*-2 mRNA is 4.6 kb long. Restri mapping of the positive clones having an insert of >1.5 kl to the identification of two distinct classes of clones (Fig.

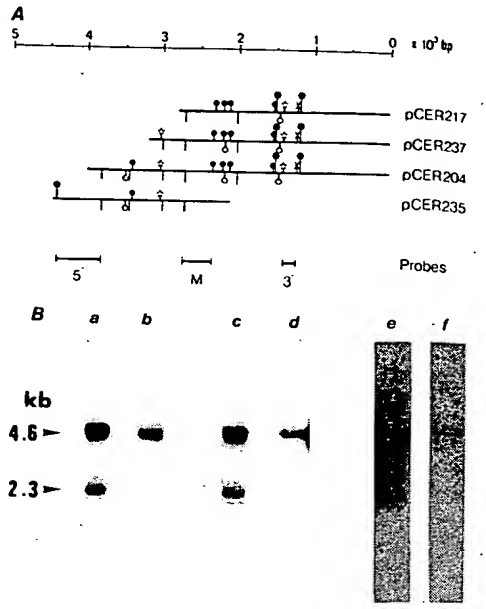

Fig. 1 Analysis of c-*erb-B*-2 clones. *A*, Restriction maps of c-*erb-B*-2 clon *B*, Northern analysis of MKN-7 and placental mRNAs. *A*, Maps w constructed by the standard procedure for restriction digestion analysis plasmid DNA[41]. Single or double digestion with restriction endonuclea *Kpn*I(×), *Eco*RI(☆), *Bam*HI (●), *Pvu*II (|), *Sma*I (*) and *Acc*I (O) w performed. The *Acc*I site at position 2.2 kb in this map is absent from t pCER217 insert but pCER237 and pCER204 were cleaved with the enzy at this position: Sequence data reveal that pCER217 carries a deletion 42 bp in this region (see Fig. 2 legend). The exon sequence of the 440- KX DNA prepared from the genomic clone λ107 (ref. 1) corresponds t 150-bp *Eco*RI-*Kpn*I fragment of pCER217 (3′ probe). A 450-bp *Pvu* *Bam*HI fragment of pCER217 (middle probe) and a 650-bp *Sma*I-*Pv* fragment of pCER235 (5′ probe) were used for the second and third scre ings, respectively, of the MKN-7 cDNA library. Total RNA was prepar from MKN-7 cells by the guanidine isothiocyanate–caesium chlori method[28]. Poly(A)$^+$ RNA was selected by two cycles of oligo(dT)-cellul column chromatography[29]. A cDNA library was constructed by the meth of Okayama and Berg[30], using 5.4 μg poly(A)$^+$ RNA and 2.8 μg vector-prin DNA. (Avian myeloblastosis virus reverse transcriptase was from Dr J. Beard.) *Escherichia coli* MC1061 (ref. 31) was used for transformation Ampicillin-resistant transformants (60,000 independent clones) were fi screened by hybridization at 60 °C for 16 h in the solution described p viously[33] with the 440-bp KX DNA fragment. Among 21 distinctive posit clones, pCER217 carried the longest insert of 2.8 kb. To obtain clo harbouring cDNA for further upstream sequence, 50,000 independent clo of the same library were screened with the 450-bp *Pvu*II–*Bam*HI fragm mapped in the 5′ portion of the pCER217 insert. Of 44 positive clon plasmid pCER237 carried the longest insert (3.2 kb). A plasmid pCER2 was assumed to be derived from an aberrant c-*erb-B*-2 mRNA which shar 5′ sequence with normal 4.6-kb mRNA, as described in the text. The MKN cDNA library (100,000 clones) was again screened with a 650-bp *Sma* *Pvu*II restriction fragment derived from pCER235; 48 positive clones w obtained, one of which, designated pCER204, contained the longest ins (4.0 kb) and the restriction maps of its 5′ one-third and 3′ two-thirds w identical to those of pCER235 and pCER217, respectively. *B*, Nitrocellul filters containing poly(A)$^+$ RNAs from MKN-7 (lanes *a*, *c*, *e*) and hum placenta (lanes *b*, *d*, *f*) were hybridized with the 5′ (lanes *a*, *b*), mid (lanes *c*, *d*) and 3′ probes (lanes *e*, *f*) shown in *A*. Lanes *b* and *d* photographs obtained after longer exposure of the film (overnight expos for lanes *a*, *c*; 2 weeks exposure for lanes *b*, *d*). Poly(A)$^+$ RNA (2 μg) fr MKN-7 cells was denatured with 50% formamide and 2.2 M formaldeh and applied to a 1% agarose gel containing 2.2 M formaldehyde[34]. RN on the gel were transferred to a nitrocellulose filter[35], which was th hybridized for 16 h under stringent conditions (50% formamide, 4×SS 42 °C) with the DNA probes. After hybridization, the filter was wash under stringent conditions as described elsewhere[1]. Under the hybridizati conditions used, no hybridization of the DNA probes with the EGF recep mRNA was observed. The DNA probes were labelled with [α-^{32}P]dC (3,000 Ci mmol^{-1}; Amersham) by nick-translation[36].

Fig. 2 The c-*erb-B*-2 cDNA nucleotide sequence and predicted amino-acid sequence. Nucleotides are numbered at both sides. Amino-acid sequence is numbered from the putative signal peptide above the sequence. Black dotted bars indicate the putative transmembrane region. The AATAAA box is followed (14 bp downstream) by the polyadenylated 3′ end of the mRNA. The nucleotide sequence from residues 1 to 1,810 was from pCER235 and that from residues 1,583 to the extreme 3′ end was from pCER217, except that the sequence of the 77-bp *Bam*HI fragment (2,314–2,390) was from pCER237, as a sequence of 42 bp was apparently deleted in pCER217 (see Fig. 1 legend). The overlapping nucleotide sequences (1,583–1,810) of pCER217 and pCER235 match. The extreme 3′ sequence of pCER235, which was derived from a sequence of unknown origin caused by chromosomal translocation is not shown here. The nucleotide sequence was determined by the Maxam-Gilbert[37] procedure and the dideoxy chain termination method in conjunction with bacteriophage M13mp19 (refs 38, 39).

suggesting that there are two species of c-*erb-B*-2 mRNA in MKN-7 cells. As cDNA clones represented by pCER235 have inserts of ≤2.3 kb, the other class of clones represented by pCER204 (or 217 and 237) was thought to be derived from the 4.6-kb mRNA. The pCER235 insert has a sequence that corresponds to the 5′ half, but not the 3′ half, of the pCER204 insert, suggesting that pCER235 was derived from mRNA which shares its sequence with the 5′ portion of normal 4.6-kb mRNA. This possibility was demonstrated by Northern hybridization of MKN-7 RNA (Fig. 1*B*). Poly(A)$^+$ RNA from MKN-7 cells was probed with DNAs specific for different portions of the cDNA clones. Using the 3′ probe, equivalent to the KX DNA probe[1], we observed a single species of 4.6-kb mRNA, which is expressed at an elevated level in MKN-7 cells (50-fold relative to placenta). However, both the *Pvu*II–*Bam*HI fragment (middle probe) and the *Sma*I–*Pvu*II fragment (5′ probe) reacted with a 2.3-kb

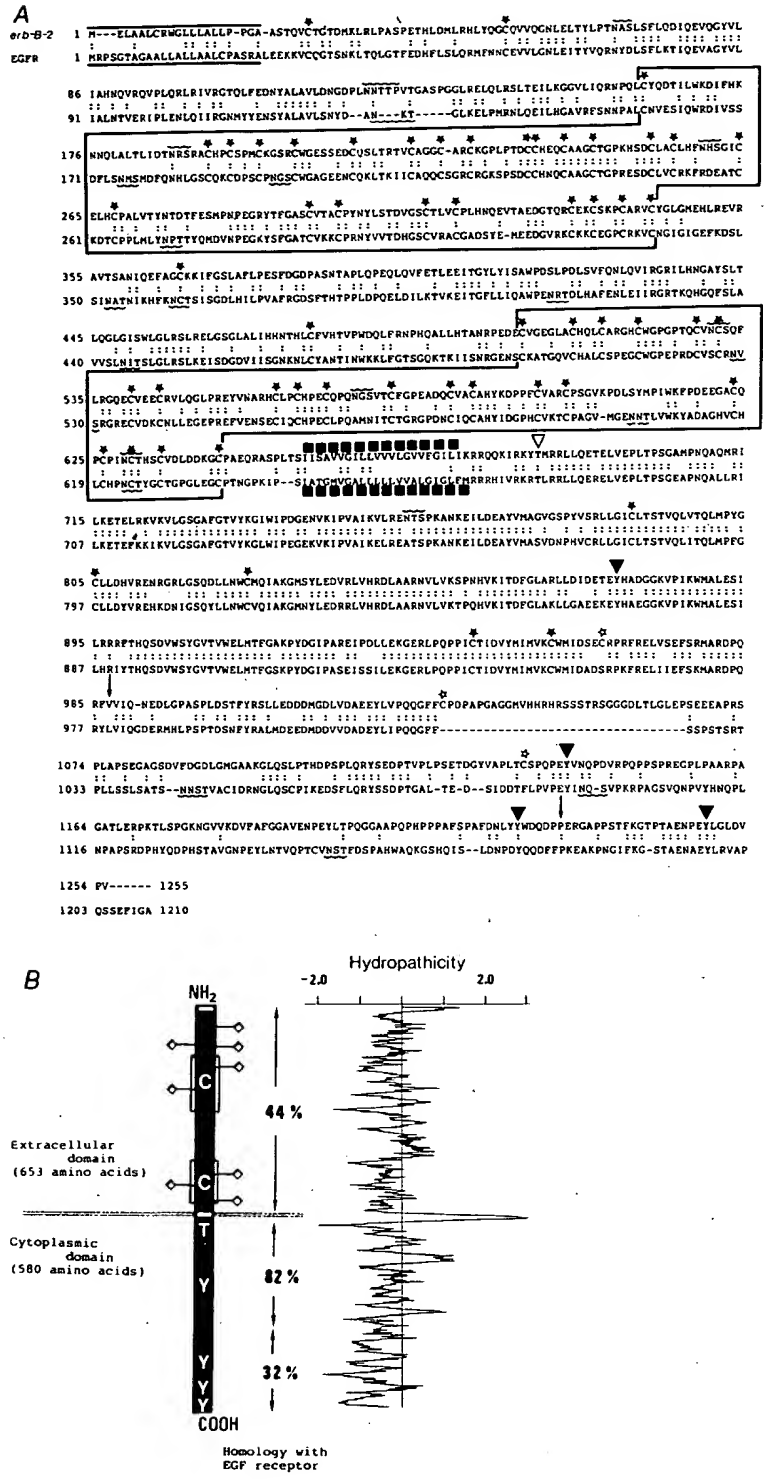

Fig. 3 Amino-acid sequence of c-*erb-B*-2. ***A***, Alignment of the amino-acid sequences of c-*erb-B*-2 and EGF receptor (EGFR). Amino acids of the two proteins are numbered on the left. Identities in the sequences are marked by two dots between the two lines; predicted transmembrane regions are represented by dotted black bars; the possible *N*-linked glycosylation sites by wavy lines; horizontal lines indicate signal peptides (putative for c-*erb-B*-2); stars indicate cysteine residues in the sequence: solid stars are common to the two proteins and the open stars are specific to c-*erb-B*-2. The major sites of threonine and tyrosine phosphorylation of the EGF receptor or pp60^{src} are conserved in c-*erb-B*-2 and are shown by open and closed triangles, respectively. The amino-acid residues corresponding to the extreme carboxy terminus of pp60^{src} and v-*erb-B* protein are indicated by vertical arrows at positions 986 and 1,228, respectively. Boxed sequences show the cysteine clusters. ***B***, Schematic illustration of the c-*erb-B*-2 protein. Amino-acid sequence homologies of the c-*erb-B*-2 protein and the EGF receptor are shown. T, threonine and Y, tyrosine are possible phosphorylation sites. C, cysteine clusters (also boxed). —◇, Possible glycosylation sites. An open box at the amino terminus shows the putative signal peptide and that in the middle shows possible transmembrane sequence. The hydropathicity of the c-*erb-B*-2 sequence is also shown.

mRNA in addition to the 4.6-kb mRNA in MKN-7 cells, but with the 4.6-kb mRNA alone in placenta. The 2.3-kb mRNA was over-produced to about the same extent as the 4.6-kb mRNA.

Thus, the nucleotide sequences of the cloned inserts of pCER235 and pCER217 (237 or 204) were assumed to represent the overall sequence of the c-*erb-B*-2 gene product. The entire nucleotide sequence of 4,480 bp obtained from pCER235, pCER237 and pCER217 is shown in Fig. 2. The longest open reading frame is composed of 3,765 nucleotides, whose translated amino-acid sequence of 1,255 residues is also shown. The predicted initiation codon ATG is flanked by nucleotides that match Kozak's criteria[4] for a translation initiation site. A primar translation product of the c-*erb-B*-2 gene was calculated to hav a relative molecular mass (M_r) of 137,895.

Previous analysis of the c-*erb-B*-2 genomic clone λ107 showe that the c-*erb-B*-2 gene product has a kinase domain that highly homologous with that of EGF receptor/v-*erb-B*[1]. As shown in Fig. 3*A*, the entire amino-acid sequences of the tw proteins are extremely similar and the hydrophilicity profile (Fig. 3*B*) and secondary structure[6] (data not shown) predicte from the amino-acid sequence of the c-*erb-B*-2 protein are al similar to those predicted for the EGF receptor. A sequence 22 amino-acid residues (654–675) is strongly hydrophobic a

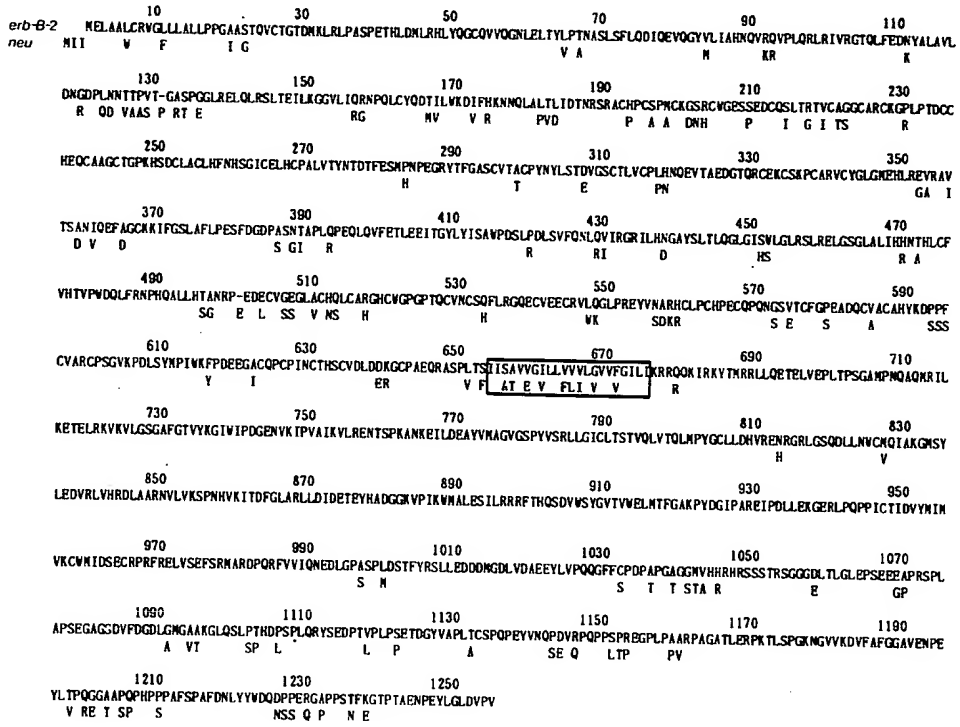

Fig. 4 Comparison of rat *neu* and human c-*erb-B*-2. The amino-acid sequence of c-*erb-B*-2 is compared with that of *neu* (kindly communicated by Dr R. A. Weinberg). Only non-identical amino acids are shown for *neu*. The amino acids for c-*erb-B*-2 are numbered above the sequence. The boxed sequence indicates the transmembrane domain.

could serve as a membrane-anchoring domain. This sequence, like those of other receptors[7-10], is followed immediately by basic amino acids (Lys-Arg-Arg), which helps in the correct allocation of the protein at the cell surface. The first 21 amino-acid residues are also highly hydrophobic, which suggests that they represent a signal sequence for membrane glycoprotein. Eight possible sites of *N*-linked glycosylation were identified in the amino-terminal moiety of the c-*erb-B*-2 protein. These data indicate that the two proteins with similar sizes have transmembrane topologies that resemble each other. Therefore, we tentatively conclude that the c-*erb-B*-2 protein is a receptor for an unknown growth factor.

The sequence of the putative extracellular domain shows 44% homology with the ligand binding domain of the EGF receptor. A striking similarity is the presence of two cysteine-rich regions, in which the spatial distribution of cysteine residues is virtually identical with that in the EGF receptor. The sequences of cysteine clusters are rather hydrophilic (Fig. 3*B*) and would facilitate the generation of a specific conformation for signal transmission through intramolecular or intermolecular S-S bridges, as is suggested for other receptors for growth factors[8-10]. These findings also suggest that the extracellular domains of the c-*erb-B*-2 protein and EGF receptor form similar configurations and bind to structurally related ligands. However, not only EGF but ligands such as tumour growth factor (TGF)-β, TGF-γ, fibroblast growth factor, erythropoietin, nerve growth factor, insulin and platelet-derived growth factor all failed to activate the kinase activity that is presumably intrinsic to the c-*erb-B*-2 protein (data not shown).

The sequence of 260 amino acids (residues 727–986), including the postulated ATP-binding site[11] of the cytoplasmic domain of the c-*erb-B*-2 protein, is homologous with the kinase domain of the EGF receptor (82% homology) and with retroviral oncogene products of the *src* family (25–40% homology[1]). Therefore, the c-*erb-B*-2 protein seems to have tyrosine kinase activity. Preliminary experiments using antiserum raised against a synthetic peptide of 14 amino-acid residues at the carboxy terminus of c-*erb-B*-2 show that the c-*erb-B*-2 protein gp185 exhibits tyrosine kinase activity (T.A. *et al.*, in preparation). The homology between the two proteins decreases to 32% in the 269 amino-acid residues at the carboxyl end. However, three major *in vitro* phosphorylation sites of the EGF receptor[12] are also conserved (tyrosine residues at 1,139, 1,222 and 1,248 of c-*erb-B*-2), indicating that the c-*erb-B*-2 protein could also be autophosphorylated.

Another interesting feature of the c-*erb-B*-2 protein is the presence of a threonine residue at position 686, surrounded by basic amino-acid residues, which is equivalent to threonine 654 of the EGF receptor on which protein kinase C-mediated phosphorylation occurs[13]. Therefore, the c-*erb-B*-2 protein may be phosphorylated by protein kinase C, which would play an important role in signal transmission as suggested for the EGF receptor.

Comparison of the nucleotide sequences and deduced amino-acid sequences of the recently characterized rat *neu* (see accompanying paper[14]) and human c-*erb-B*-2 (our present results) reveal that the *neu* gene is the rat counterpart of the c-*erb-B*-2 gene (see Fig. 4 for the amino-acid sequences). Surprisingly, only two amino acids in the kinase domain (residues 813 and 817) differ between the two proteins, although we do not know whether the c-*erb-B*-2 gene of MKN-7 cells can transform NIH 3T3 cells. The possible glycosylation sites of the *neu* product are located at positions corresponding to those of the c-*erb-B*-2 product, except that no sequence corresponding to Asn-Asn-Thr-Thr (positions 124–127 of c-*erb-B*-2) is found in *neu*, suggesting that the mature c-*erb-B*-2 protein is as large as the *neu* gene product gp185.

Evidence is accumulating that amplification and overexpression of a proto-oncogene can cause cell transformation *in vitro* and can play a part in the neoplastic process of human tumours[15-20]. Amplification and elevated expression of the EGF receptor gene have been observed in glioblastomas and in squamous carcinoma cell lines[21-24]. In contrast, amplification of the c-*erb-B*-2 gene has been seen in three human adenocarcinomas: one salivary adenocarcinoma[1], one mammary carcinoma[2] and one MKN-7 gastric cancer cell line, suggesting that increased expression of the c-*erb-B*-2 gene provides a selective advantage in the formation or the progress of tumours of epithelial cells.

Because the cDNA clone pCER235 derived from the 2.3-kb

mRNA does not contain a signal for membrane anchoring of the predicted polypeptide (data not shown), the translation product of the 2.3-kb mRNA, corresponding to the sequence of the c-*erb*-*B*-2 ligand-binding domain, should be secreted by MKN-7 cells. Similarly, a truncated EGF receptor sequence has been reported to be over-expressed concomitantly with the EGF receptor in A431 cells[7,25,26], as a consequence of a strong aberration of the chromosomes[27]. Production of the truncated receptor in MKN-7 cells may also be caused by chromosomal aberration[40]. We cannot exclude the possibility that extracellular accumulation of truncated growth factor receptors is associated with the appearance of the transformed phenotype. It is also possible that they provide a growth advantage to the cells in culture, in which the growth factor receptors are over-expressed.

We thank N. Kitamura, H. Okubo and S. Nakanishi for help in constructing the MKN-7 cDNA library, H. Kawano for technical assistance and S. Sasaki for help in preparing the manuscript. We also thank Y. Kaziro for critical reading of the manuscript.

Received 28 August; accepted 27 November 1985.

1. Semba, K., Kamata, N., Toyoshima, K. & Yamamoto, T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6497-6501 (1985).
2. King, C. R., Kraus, M. H. & Aaronson, S. A. *Science* **229**, 974-976 (1985).
3. Schubert, D. *et al. Nature* **249**, 224-227 (1974).
4. Kozak, M. *Nucleic Acids Res.* **9**, 5233-5252 (1981).
5. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105--132 (1982).
6. Chou, P. Y. & Fasman, G. D. *A. Rev. Biochem.* **47**, 251-276 (1978).
7. Ullrich, A. *et al. Nature* **309**, 418-425 (1984).
8. Ebina, Y. *et al. Cell* **40**, 747-758 (1985).
9. Ullrich, A. *et al. Nature* **313**, 756-761 (1985).
10. Yamamoto, T. *et al. Cell* **39**, 27-38 (1984).
11. Kamps, M. P., Taylor, S. S. & Sefton, B. M. *Nature* **310**, 589-592 (1984).
12. Downward, J., Parker, P. & Waterfield, M. D. *Nature* **311**, 483-485 (1984).
13. Hunter, T., Ling, N. & Cooper, J. A. *Nature* **311**, 480-483 (1984).
14. Bargmann, C. I., Hung, M.-C. & Weinberg, R. A. *Nature* **319**, 226-230 (1986).
15. Chang, E. H., Furth, M. E., Scolnick, E. M. & Lowy, D. R. *Nature* **297**, 479-483 (1982).
16. Gazit, A. *et al. Cell* **39**, 89-97 (1984).
17. Clarke, M. F. *et al. Nature* **308**, 464-467 (1984).
18. Schwab, M. *et al. Nature* **305**, 245-248 (1983).
19. Kohl,, N. E. *et al. Cell* **35**, 359-367 (1983).
20. Schwab, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4940-4944 (1984).
21. Libermann, T. A. *et al. Cancer Res.* **44**, 753-760 (1984).
22. Libermann, T. A. *et al. Nature* **313**, 144-147 (1984).
23. Cowley, G., Smith, J. A., Gusterson, B., Hendler, F. & Ozanne, B. in *Cancer Cell 1* (eds Levine, A. J., VandeWoude, G. F., Topp, W. C. & Watson, J. D.) 5-10 (Cold Spring Harbor Laboratory, New York, 1984).
24. Yamamoto, T. *et al. Cancer Res.* (in the press).
25. Merlino, G. T. *et al. Science* **224**, 417-419 (1984).
26. Weber, W., Gill, G. N. & Spiess, J. *Science* **224**, 294-297 (1984).
27. Merlino, G. T. *et al. Molec. cell. Biol.* **5**, 1722-1734 (1985).
28. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
29. Aviv, H. & Leader, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408--1412 (1972).
30. Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).
31. Casadaban, M. & Cohen, S. N. *J. molec. Biol.* **138**, 179-207 (1980).
32. Hanahan, D. & Meselson, M. *Meth. Enzym.* **100**, 333-342 (1983).
33. Iwai, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7405-7409 (1983).
34. Lehrach, H., Diamond, K., Wozney, J. M. & Boedtker, H. *Biochemistry* **16**, 4743-4751 (1977).
35. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
36. Maniatis, T., Jeffery, A. & Kleid, D. *Proc. natn. Acad. Sci. U.S.A.* **12**, 1184-1188 (1975).
37. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 109-111 (1979).
38. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
39. Messing, J. *Meth. Enzym.* **101**, 20-77 (1983).
40. Fukushige, S. *et al. Molec. cell. Biol.* (in the press).
41. Maniatis, T., Fritsch, E. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Primate η-globin DNA sequences and man's place among the great apes

Ben F. Koop*, Morris Goodman*§, Peilin Xu†, Keith Chan† & Jerry L. Slightom†‡

*Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201, USA
†Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA
‡Upjohn Company, Kalamazoo, Michigan 49001, USA

Molecular studies indicate that chimpanzee and gorilla are the closest relatives of man (refs 1–7 and refs therein). The small molecular distances found point to late ancestral separations[2,4,7], with the most recent being between chimpanzee and man, as judged by DNA hybridization[3,8]. Kluge[9] and Schwartz[10] contest these conclusions: morphological characters group a chimpanzee–gorilla clade with the Asian ape orang-utan in Kluge's cladistic study and with an orang-utan–human clade in Schwartz's study. Clearly, extensive sequencing of nuclear DNA is needed to resolve by cladistic analysis the branching order within Hominoidea[11]. Towards this goal, we are sequencing orthologues of the primate $\psi\eta$-globin locus[12,13]. Here, we compare the newly completed sequences of orang-utan and rhesus monkey with human, chimpanzee, gorilla, owl monkey, lemur and goat orthologues. Our findings substantially increase the evidence indicative of a human–chimpanzee–gorilla clade with ancestral separations around 8 to 6 Myr ago. We also verify that neutral hominoid DNA evolved at markedly retarded rates.

The η locus is one of five ancient β-related globin genes linked in a cluster 5′-ε-γ-η-δ-β-3′ that arose from tandem duplications (200–100 Myr)[12,13]. This ancient η gene was embryonically expressed in early eutherians and persisted as a functional gene in artiodactyls, but became a pseudogene in proto-primates and was lost from rodents and lagomorphs. Previous work from this laboratory[13] established that the goat η gene sequence

§To whom correspondence should be addressed.

Fig. 1 (Opposite) Aligned nucleotide sequences of seven primate $\psi\eta$-globin genes and the active goat η-globin gene. Total DNA was isolated from orang-utan (*Pongo pygmaeus*) no. 1 liver (Yerkes Primate Center) and from a rhesus monkey (*Macaca mulatta*) blood sample (California Primate Center). DNAs were subjected to a series of limited *Eco*RI digestions and size-selected fragments, 15–25 kilobases (kb), were cloned into the λ vector Charon 32[33] by the procedure described by Slightom *et al.*[34]. Recombinant phage DNAs were packaged into phage capsids using the *in vitro* phage packaging procedure[35]. Charon 32 phage were plated on the recA⁻ *Escherichia coli* host ED8767[36] and screened using a ^{32}P-labelled 245-base-pair (bp) *Ava*II–*Eco*RI fragment isolated from the γ-globin cDNA clone pJW151[37]. The $\psi\eta$-containing *Eco*RI fragments were subcloned into pBR322 (orang 7.0 kb, rhesus 10.0 kb). DNA sequencing was done using the chemical procedure described by Maxam and Gilbert[38]. The nucleotide sequences for human, chimpanzee and gorilla are from Chang and Slightom[16], and owl monkey and lemur are from Harris *et al.*[12]. Only the 5′-flanking to exon 2 of the lemur sequence is orthologous to the η gene[12,15]; therefore only this part of the sequence is used in our analyses. The goat η-globin gene (goat ϵ^{11} gene) nucleotide sequence is from Shapiro *et al.*[14]. The nucleotide sequencing numbering system is based on the overall alignment among these sequences. The complete nucleotide sequence for the human $\psi\eta$-globin gene (HUMA) is presented on the top line and differences are given for the remaining sequences. Asterisks indicate the presence of gaps placed to minimize the number of geneic changes during descent[13,39]. The simian $\psi\eta$ genes are divided into three exons separated by two introns each obeying the GT-AG splicing rule[40]. Lemur $\psi\eta^{12,15}$ extends only to the 3′ end of exon 2 and has a defective intron 1 splice site. Intron 1 sequences are all 121 bp in length, while intron 2 lengths vary between 841 to 877 bp. The 5′-flanking region of the orang $\psi\eta$ gene contains a 38-bp direct repeat (positions 205–244 and 245–282) common to hominoid $\psi\eta$ genes. Each repeat contains a CCAAT type promoter element. The RNA polymerase II binding site (TATAA) has diverged in the $\psi\eta$ genes from that found in normal β-type globin genes[41] (note positions 299–304). As in the other primate $\psi\eta$ genes the initiation codon (INT positions 384–386) of orang and rhesus shows defects that prohibit translation. All of the primate $\psi\eta$ genes share a terminator sequence TGA (positions 1,885–1,888) and have the canonical poly(A) addition signal[42]. With regard to the direct repeat at position 1,895–1,914 and 1,915–1,940, the 22-base gap (positions 1,895–1,916) in orang and rhesus could be the result of either two independent deletions or a deletion of the 5′ direct repeat in the stem catarrhine followed by a reduplication in the stem hominines. Exons terminate at the ⟶ sites.